DOCUMENT RESUME

ED 464 913 TM 034 157

AUTHOR	Shaftel, Julia; Yang, Xiangdong
TITLE	Special Education Participation in Statewide Accountability Assessment: Analysis of Options Using Item Response Theory.
PUB DATE	2002-04-00
NOTE	25p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 1-5, 2002).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Academic Accommodations (Disabilities); *Accountability; Curriculum; *Disabilities; *Elementary School Students; Intermediate Grades; *Item Response Theory; *Special Education; Test Format; Test Items

ABSTRACT

A group of fourth-grade students with disabilities participating in statewide mathematics assessments in the spring of 2001 (n=570) was given a new assessment focused on specifically modified state curricular standards defined for students with disabilities. A block of items from that assessment was also given to students taking the general mathematics assessment (n=1,944, including 235 students with disabilities), allowing the difficulty of all items to be estimated on a common scale using item response theory methodology. This study compared the characteristics of the two tests. Findings support the conclusion that a simpler test, based on the same curricular objectives, but providing better measurement of low-achieving students, could be developed and used for this special population. (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 464 913

Special Education Participation in Statewide Accountability Assessment: Analysis of Options Using Item Response Theory

American Educational Research Association

National Meeting

April 1, 2002

Julia Shaftel and Xiangdong Yang

Center for Educational Testing and Evaluation

University of Kansas

BEST COPY AVAILABLE

TM034157

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
J. Shaftel
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Abstract

A group of fourth grade students with disabilities participating in statewide mathematics assessments in the spring of 2001 was offered a new assessment based on specifically modified state curricular standards. A block of items from that assessment was also given to students taking the general mathematics assessment, allowing the difficulty of all items to be estimated on a common scale using item response theory methodology. This project compared the characteristics of the two tests and supports the conclusion that a simpler test, based on the same curricular objectives but providing better measurement of lower-achieving students, could be developed and used for this special population.

Special Education Participation in Statewide Accountability Assessment:

Analysis of Options Using Item Response Theory

The 1997 amendments to the Individuals with Disabilities Education Act require that all students participate in statewide accountability assessments, including students with all types and severity of disability. The majority of students with disabilities should participate in general assessments because this promotes greater instructional opportunity and higher achievement expectations for students who have historically been exempted from accountability testing and hence from measurement of their learning (Thurlow, Elliott & Ysseldyke, 1998). However, general assessments may not be valid for all students with disabilities because of the lack of correspondence between appropriate instruction and items on the test. Therefore, in addition to general curricular assessments, alternative forms of tests as well as a variety of accommodations and modifications have been proposed and developed in order to include all students with disabilities in accountability testing.

During the spring of 2001, students in fourth, fifth, sixth, seventh, eighth, tenth and eleventh grades took statewide accountability assessments, and all students, including those with disabilities, were expected to participate. In Kansas, an Alternate Assessment was designed for students with the most significant disabilities consisting of a portfolio of items demonstrating the student's performance during the year and a rating scale completed during interviews with parents and teachers in the spring. The eligibility criteria for this assessment were intentionally quite restrictive. In order to qualify for participation in this assessment, a student must be instructed in a curriculum that corresponds to the state's Extended Curricular Standards in Reading, Writing and Mathematics, which are downward extensions of the state's general

curricular standards. Students participating in the Alternate Assessment during 2001, the first year of implementation, comprised only about 0.75% of assessed students.

Because the constraints on participation in the Alternate Assessment resulted in such a small proportion of eligible students, a pool of students with disabilities remained for whom the general assessment was still too difficult, did not correspond well with their adapted curricular needs, and hence was not a valid measure of their skills. Our state chose to develop a third option for this "gray area" of students, a series of assessments based on general curricular standards but with specific modifications built in. Because these assessments differed in substantial ways from the general assessments, they are considered to be alternate assessment options as defined by the IDEA 1997 amendments. These assessments were offered in reading, mathematics, science and social studies, and the fourth grade mathematics assessment was selected for more intensive study and comparison to the general assessment.

Accommodations v. Modifications

Accommodations have been defined as those alterations to test presentation, setting, timing, scheduling, and response that mitigate the barrier of disability and allow a student with disabilities to demonstrate actual achievement in a particular academic area, without changing the underlying construct of what is being measured (Hollenbeck, Tindal & Almond, 1998; Schulte, Elliott, & Kratochwill, 2001; Thurlow, Elliott & Ysseldyke, 1998). Examples include eyeglasses or large print for visually impaired students, frequent breaks or an isolated setting for students with attention problems, and allowing students with writing or motor difficulties to dictate answers or mark responses in the test booklet rather than on the answer sheet. Appropriate accommodations are those used regularly for instruction and tailored to individual learning needs. Ideally, accommodations selectively benefit students with special needs without

conferring an undue advantage; students without those special needs would not experience a benefit from the accommodation (Hollenbeck, Tindal & Almond, 1998; Schulte, Elliott, & Kratochwill, 2001). For this reason, some accommodations used frequently for instruction, such as oral presentation of reading passages, paraphrasing, or using a calculator for computation items, cannot be permitted during testing without calling into question the meaning of the constructs measured by the test. Because of the controversial nature of some accommodations and their unknown impact on test score comparability, the selection and use of accommodations for special needs populations is currently the topic of a great deal of research (Destefano, Shriner & Lloyd, 2001; Johnson, Kimball, Brown & Anderson, 2001; Schulte, Elliott, & Kratochwill, 2001).

Accommodations that are likely to change the nature of what is being tested have been called modifications in order to distinguish them from the better-understood accommodations that are believed to preserve score comparability (Hollenbeck, Tindal & Almond, 1998; McDonnell, McLaughlin & Morison, 1997; Schulte, Elliott, & Kratochwill, 2001). Modifications are assumed to change the test content and may include deleting items or response options or changing open-ended, constructed response items into multiple choice items (Elliott & McKevitt, 2000), reading items intended to measure reading comprehension aloud, paraphrasing or simplifying questions, or permitting the use of dictionaries or calculators.

In Kansas the terms accommodation and modification are not used interchangeably. A clearly defined set of accommodations is available to any student depending on individual need and regular instructional use, not on disability status or label. These accommodations are reported on a student-by-student basis for each assessment. Modifications that may change the nature of the test and nullify score comparability are not permitted except in certain

circumstances. For example, calculator use on mathematics assessments is not permitted at fourth grade except for students with disabilities who have that modification noted in their Individualized Education Plans (IEPs) or 504 plans, and only a handful of students actually accessed that accommodation. Some modifications, such as oral presentation of reading passages, are prohibited for all students.

Even though a number of accommodations are available, these may not provide sufficient access to the general instructional curriculum or corresponding assessments for some lower-performing students with disabilities. This is particularly true in a state such as ours in which accountability assessments are designed to be rigorous, grade-level assessments over a broad range of curricular content rather than minimum competency tests. In order to include all students in statewide assessments and still attempt to obtain instructionally relevant results rather than random responses, assessments with specific and controlled modifications were developed in anticipation of the IDEA 1997 requirements that went into effect in July 2000. One of these was the mathematics assessment with modifications designed for fourth grade students with disabilities who were not eligible for the portfolio/rating scale Alternate Assessment.

A major goal in developing the modified assessments was to control and regularize the types of modifications that students with disabilities may need for assessment by building them into the test itself. Test modifications, like accommodations, should parallel regular instructional modifications that are typical for students with disabilities. Including the modifications within the standardized test protocol should minimize the validity problems of allowing modifications to be developed locally by IEP teams, which may then result in an assessment given in myriad non-standardized ways for which scores can no longer be meaningfully compared. The modifications permitted on this assessment were carefully distinguished from accommodations.

such as extra time, frequent breaks, oral presentation of non-reading items, and dictation of answers, that, if regularly used for instruction, are available to any student on any assessment, including the modified assessments. This research evaluated the construction and use of this test with the following general questions in mind: How does the new test compare to the general assessment instrument? Can an assessment with built-in curricular modifications be an effective tool for evaluating the achievement of lower-performing students with disabilities in the general curriculum? For which students with disabilities is the new modified assessment a better measure of mathematics skills?

Method

Modified Assessment Instrument

Items for the modified mathematics test were written to measure instructional objectives, called indicators, defined by the state for students with disabilities and including all of the conceptual content of the general curricular indicators. Modifications were made to the existing mathematics curricular indicators at each grade level (fourth, seventh and tenth grades) by a team of mathematics and special education teachers from that grade level. These teams were directed by the coordinator of mathematics curriculum for the Kansas State Department of Education.

The indicators were modified from the general curricular indicators in a variety of ways, including:

1) simplifying operands, such as restricting computation to whole numbers instead of decimals or limiting the number of decimal places to be manipulated;
2) limiting the number of steps or operations to be performed;

3) limiting abstract content by requiring that items be contextually relevant to students with disabilities.

In addition to indicator modifications, test and item modifications were also defined, including

4) removing extraneous information from word problems;
5) simplifying the language or context of the problem, for example, reducing a Venn diagram to two attributes instead of three;
6) using additional illustrations or graphics, such as showing pictures of items with price tags to illustrate a money estimation problem;
7) putting fewer items on a page;
8) including key definitions and formulas within the problem;
9) permitting calculators throughout the assessment;
10) reducing the total number of items presented to the student.

Examples of original and modified indicators are shown in Table 1.

The resulting assessment at the fourth grade level consisted of 35 items rather than the 52 items in the general test. Several presentation and response accommodations were made available to all students who participated in this modified assessment. For example, students were expected to be tested in smaller groups, most likely by their special education teachers, rather than as part of a general mathematics class. Teachers were expected to provide additional help with filling out answer sheets and marking answers, such as transcribing answers to bubble sheets for students who marked their answers in test booklets, and students were to have as much time as they needed to thoughtfully and carefully complete each day's set of problems.

Participants

For this first year of implementation, no clear guidelines were available to assist IEP teams in determining which students would be best served by the modified test, so IEP teams made the initial test selection decisions on a student-by-student basis. Teams were given the restriction from the state department of education that students with disabilities must score below the 2.5th percentile on a norm-referenced standardized test of mathematics in order to be eligible for the modified mathematics test. However, the modified test's characteristics were not known and it was unclear whether that guidance would match students to the best assessment option. Furthermore, it was not known what overall percentage of students would turn out to be eligible for the modified assessment using existing test data from their files. The state wanted to balance the two objectives of keeping that number to a minimum while still providing the most valid assessment to each student with disabilities.

A total of 570 students were deemed eligible for the modified assessment by their IEP teams and took the 35-item modified assessment as their only math achievement test. In addition to students with disabilities taking the modified assessment, 1944 fourth graders taking the 52-item general assessment, including 235 students with disabilities, were randomly sampled by whole classes from across the state. Fifteen items from the modified assessment were assembled into a test booklet and administered to these students as an additional test section. These students completed the extra test booklet before or after the four regular test sessions of the general assessment.

The general assessment sample consisted of 963 females and 979 males while 232 females and 338 males comprised the modified test group. Twenty-one percent of the students taking the modified test were African-American compared to 8.5% of students in the general

assessment, and 56% of the modified test group was white compared to 70% of the general assessment sample. These gender and ethnic differences are typically found in disability populations, which tend to contain relatively more males and minority students than the general population.

Analysis

A one-parameter item response theory (IRT) model was used to place all 87 items from both the general and modified assessments onto the same scale and compute item difficulty parameters. Mathematics ability levels for all 2514 students were then estimated with the general assessment group as the reference group. Item difficulty and student ability were both scaled to have a mean of 0 and a standard deviation of 1 so that direct comparison between items and student abilities could be made. Test information functions were computed in order to determine which test would provide the most information about the achievement of students of different abilities. Test information functions were computed by summing the amount of information each item in the test provided at each ability level, then plotting those sums across the range of abilities.

Results

Item difficulties for the two tests were examined and compared, as shown in Figure 1. Item difficulties ranged from –3.6 to 2.1, with a range of –3.4 to 2.1 and a mean of –0.51 on the general test. The modified test items ranged from about the middle to slightly below the easiest items on the general test, from –3.6 to –0.5 with a mean of –1.73, demonstrating that the modified test was of lower difficulty overall. As reliabilities were good for both tests, (general coefficient α = .87, modified coefficient α = .85), this confirmed that a major objective of the

modified test had been accomplished, which was to provide lower level items and omit very difficult items while maintaining the curricular coverage of the general mathematics assessment.

The distribution of student mathematics ability is shown in Figure 2. Ability level scores for students with and without disabilities who took the general assessment ranged from –2.86 to 2.58. Since they were the scaling reference group their mean score was preset at 0 with a standard deviation of 1. Scores for students who took the modified test, all of whom had disabilities, ranged from –3.6 to 2.88 with a mean of –1.81. Six students with disabilities were assigned spuriously high ability scores on the basis of having answered 33 or more of the 35 items on the modified test correctly. Their ability might have been more accurately assessed with the general test, illustrating the importance of an assessment with sufficient items at the appropriate level of difficulty.

Since the two distributions of item difficulty and student ability use the same metric in IRT analysis, student ability can be directly compared with item difficulty. Modified test items, with difficulties ranging from –3.6 to –0.5, correspond quite closely to the range of mathematics ability evidenced by most of the students who were deemed eligible for that test by their IEP teams (omitting the six incorrectly assigned students), ranging from –3.6 to 0.69 with a mean of –1.81.

The question of which test is more appropriate for which students was addressed by computing test information curves for the two tests. The test information curves shown in Figure 3 consist of the sums of item information (pq, where p = probability of passing the item at that ability level and $q = 1-p$) for all items on a test at each ability level. Maximum item information is available when an individual has a 50/50 chance of passing the item ($pq = .25$); progressively less information is available as the item becomes too hard or too easy for that ability level. The

range of appropriate ability levels assessed by each test is apparent, confirming the visual evidence from Figures 1 and 2. Student ability ranges can be compared with the test information curves to show that the modified test provides more information about lower performing students.

The point at which one test provides more information than the other is the ability level at the crossover point of the two test information curves in Figure 3. This level can be pinpointed with the relative efficiency function, the ratio of the information provided by the modified test to the information provided by the general test. The crossover point occurs at the ability level at which the tests provide equal information. The relative efficiency function shows that students with a mathematics ability of less than –1.14 are more reliably assessed with the modified test.

The ability distribution from the general assessment sample was used to assess the cutoff point at which the modified assessment provides more test information because that distribution represents the distribution of mathematics ability across the state. Using that distribution, the math ability levels that fell below -1.14 and thus within the range of the modified test would include approximately 11.4% of the general assessment sample of students.

Discussion

These results demonstrate that a test measuring the general curriculum with reduced difficulty could be developed for students with disabilities. The modified test samples the lower range of difficulty with more items than the general assessment and omits items in the upper half of the difficulty range that would provide little additional achievement information about students whose disabilities make them eligible for the modified assessment in the first place. This was accomplished while measuring the same curricular standards and conceptual content as

the general test with pre-determined and standardized modifications built in. This more accurate measurement was achieved even though the modified test was restricted to fewer items with a table of content specifications identical to the general test.

The results also provide an estimate of the proportion of lower-achieving students for whom the modified test might be a more appropriate assessment. In this study, the lowest 11.4% of the general assessment sample distribution of math ability could be measured equally well or better by the modified test. It should be emphasized that such a high proportion of students is far too high to be instructed with a modified curriculum and diverted from the general assessment. That a relatively large proportion of students fall in this range is directly due to the rigor of the general assessment in Kansas, which contains items of considerable difficulty even for non-disabled students. This proportion does not represent the number of students who *should* take the modified test instead of the general assessment. It merely shows that the simpler modified test provides, in terms of items at the appropriate difficulty, more information about the performance of students at this level of mathematics achievement. These estimates are merely meant to compare these two assessment instruments; they do not address the relationship of the student's instructional curriculum to the items on either assessment, which is the charge of the student's IEP team when making assessment decisions. In this state, however, IEP teams can be reassured that their lower-achieving 4th grade students with disabilities were indeed measured more accurately with the modified test.

The alignment of instructional curriculum with test content is a critical assessment issue in these days of ever-increasing use of standardized tests. The IDEA 1997 amendments, in mandating that students with special needs be included in district- and state-wide testing, were driving toward just this type of alignment with the goal that special-needs students be exposed to

the general curriculum, instead of the more limited instruction that may occur in special education classrooms, in order to compete on large scale assessments. Hence diverting these same students from the assessments that are intended to improve their instructional opportunities is counter to the spirit of the federal law.

The tension between accessibility and lofty standards is high when assessments are deliberately *not* "dumbed down" in order to allow most students to achieve passing scores but are intentionally rigorous and broad in scope. High standards result in tests on which few students obtain outstanding scores and many achieve only minimal or basic proficiency, which may be an accurate representation of student performance levels. In this situation, however, students with disabilities who perform within the lowest few percentile ranks of the achievement distribution are truly not being fairly assessed on the general assessments. In order to be reasonably assessed in a manner that dignifies their learning achievements and provides real information about their progress to their schools and districts, alternate forms of assessment such as the modified test studied in this research must be developed, utilized and evaluated.

A final conclusion of this research was that IEP teams overwhelmingly made correct choices when they assigned students with disabilities to the modified rather than the general assessment. The six students who scored too high on the modified assessment represent about 1% of this special-needs population who perhaps should have been assigned to the general assessment, while another 5% who scored at the floor or chance level on the modified test might have been better evaluated using the individualized portfolio/rating scale Alternate Assessment. This leaves about 94% of students with disabilities in the "gray area" between the Alternate Assessment and the general math test for whom IEP teams made the correct test selection, even

though it was the first year of mandated participation and the precise nature of the available options was unknown.

Limitations of this study must be noted. The results are applicable only to the two specific assessments devised for 4th grade students in this state; no generalization to other instruments, populations or content areas can be made. The project was intended to evaluate whether a modified instrument could be devised that matched the content of the general test in a manner that would allow lower-achieving students to demonstrate their knowledge of mathematics content and to provide information about which students could be diverted to that test. Further study is needed with other populations, other grade levels, and other content areas to determine whether assessment options meeting the IDEA 1997 guidelines, or newer mandates currently being considered, can be developed on a broader scale. Such assessments could fulfill both the letter and the spirit of federal law requiring inclusion of all students in assessments while providing valid measurement of student progress within the general curriculum.

References

Destefano, L., Shriner, J. G., & Lloyd, C. A. (2001). Teacher decision making in participation of students with disabilities in large-scale assessment. Exceptional Children, 68, 7-22.

Elliott, S. N., & McKevitt, B. C. (2000, April). Testing accommodations decisions: Legal and technical issues challenging educators. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA. Available: http://www.wcer.wisc.edu/testacc/publications/AERApaper4-2000.htm

Hollenbeck, K., Tindal, G., & Almond, P. (1998). Teacher's knowledge of accommodations as a validity issue in high-stakes testing. Journal of Special Education, 32, 175-183.

Johnson, E., Kimball, K., Brown, S. O., & Anderson, D. (2001). A statewide review of the use of accommodations in large-scale, high-stakes assessments. Exceptional Children, 67, 251-264.

McDonnell, L. M., McLaughlin, M. J., & Morison, P. (Eds.). (1997). Educating one and all: Students with disabilities and standards-based reform. Washington, D.C.: National Academy Press.

Schulte, A. A. G., Elliott, S. N., Kratochwill, T. R. (2001). Effects of testing accommodations on standardized mathematics test scores: An experimental analysis of the performances of students with and without disabilities. School Psychology Review, 30, 527-547.

Thurlow, M. L., Elliott, J. L., & Ysseldyke, J. E. (1998). Testing students with disabilities: Practical strategies for complying with district and state requirements. Thousand Oaks, CA: Corwin Press.

17

Table 1

Original and modified mathematics indicators

Mathematics content area	Original indicator	Modified indicator
Number and computation: number sense	Determines reasonableness of numerical values involving whole numbers to 1,000,000, simple fractions, and decimals to the thousandths.	Determines reasonableness of numerical values involving whole numbers to 1,000,000, simple fractions, and decimals to the hundredths.
Number and computation: computation	Performs whole number division using dividends with up to three digits and a one-digit divisor.	Performs whole number division without remainders using dividends with up to three digits and a one-digit divisor.
Algebra: variables, equations and inequalities	Formulates and solves problem situations involving one-step equations in one unknown with a whole number solution.	Solves one-step equations involving one unknown with a whole number solution such as finding any missing number in a multiplication or division equation based on the multiplication and division facts for numbers up to 12 times 12, equations involving money such as 8 quarters + 10 dimes = □ dollars and 100 x Δ = 600. Geometric figures such as a square or

		triangle will be used to represent the unknown
Geometry: measurement and estimation	Formulates and solves real-world problems by applying measurements and measurement formulas.	Solves real-world problems by applying measurements and measurement formulas.
Data: statistics	Uses data analysis to make reasonable inferences, decisions/predictions, and to develop convincing arguments from data displayed in a variety of formats: *frequency tables* *horizontal and vertical bar graphs* *Venn diagrams or other pictorial displays* *charts and tables* *line graphs* *pictographs*	Uses data analysis to make accurate inferences from data displayed in a variety of formats: *frequency tables* *horizontal and vertical bar graphs* *Venn diagrams for up to two attributes* *charts* *pictographs*

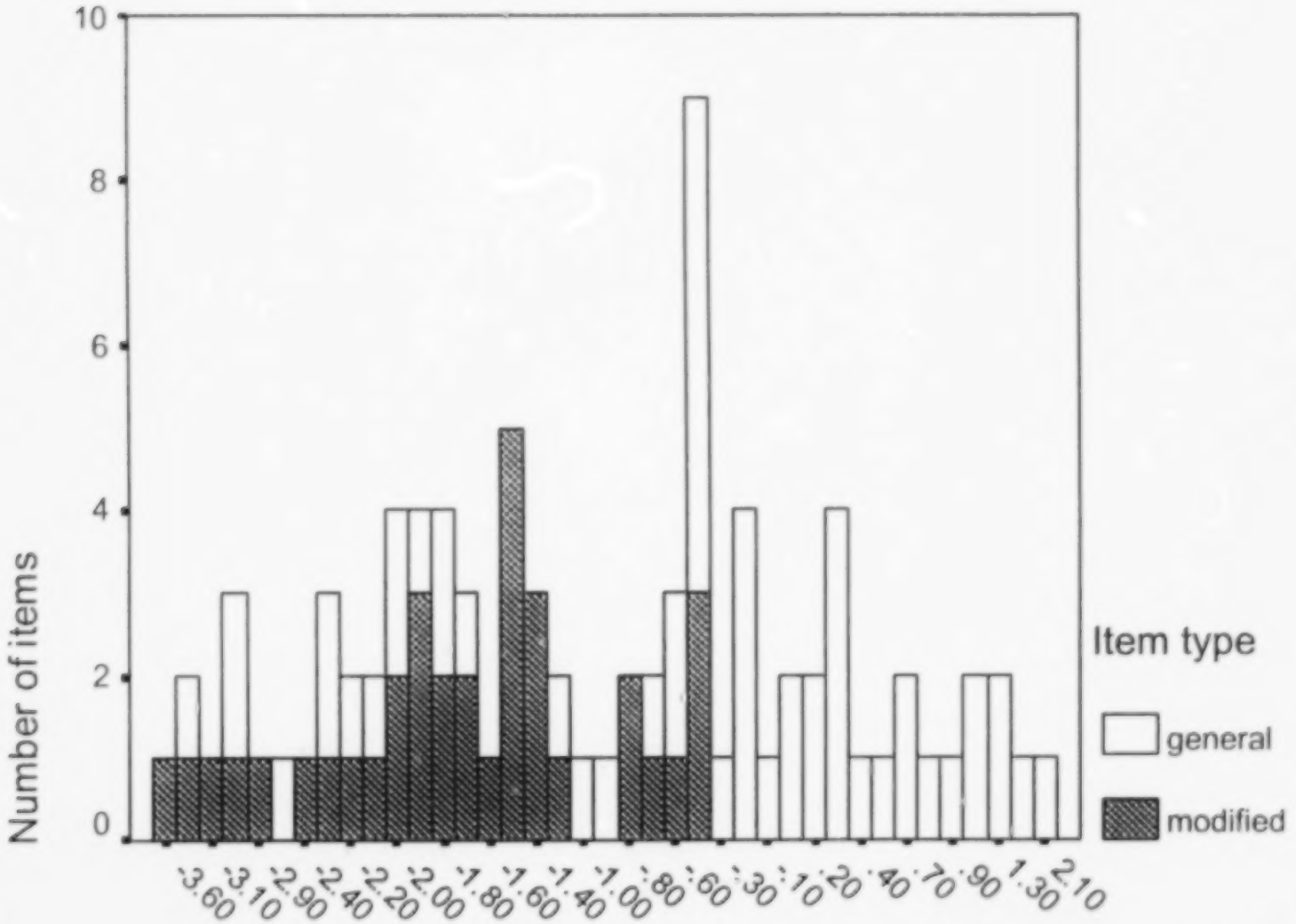

Figure 1. Item difficulty

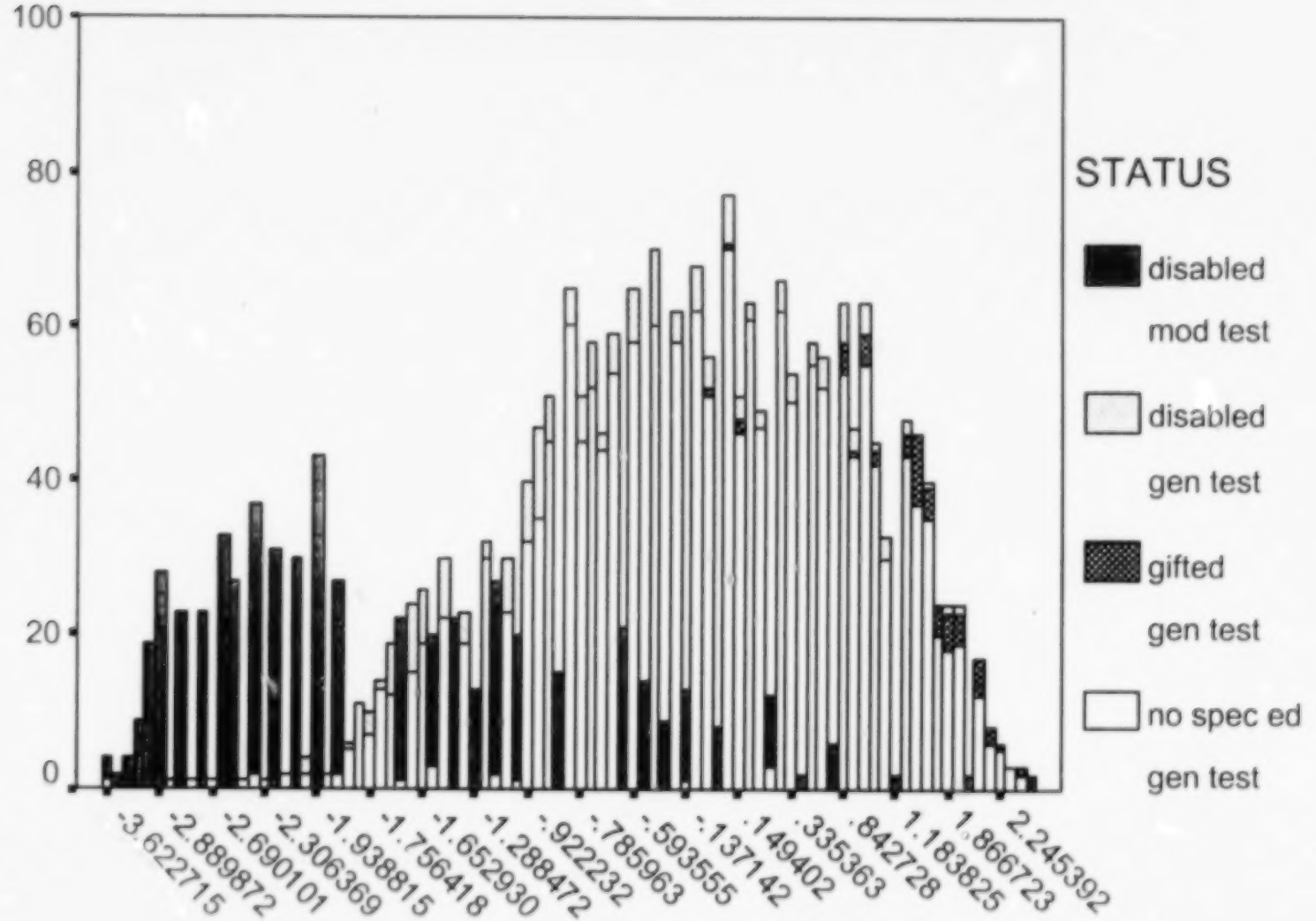

Figure 2. Math ability

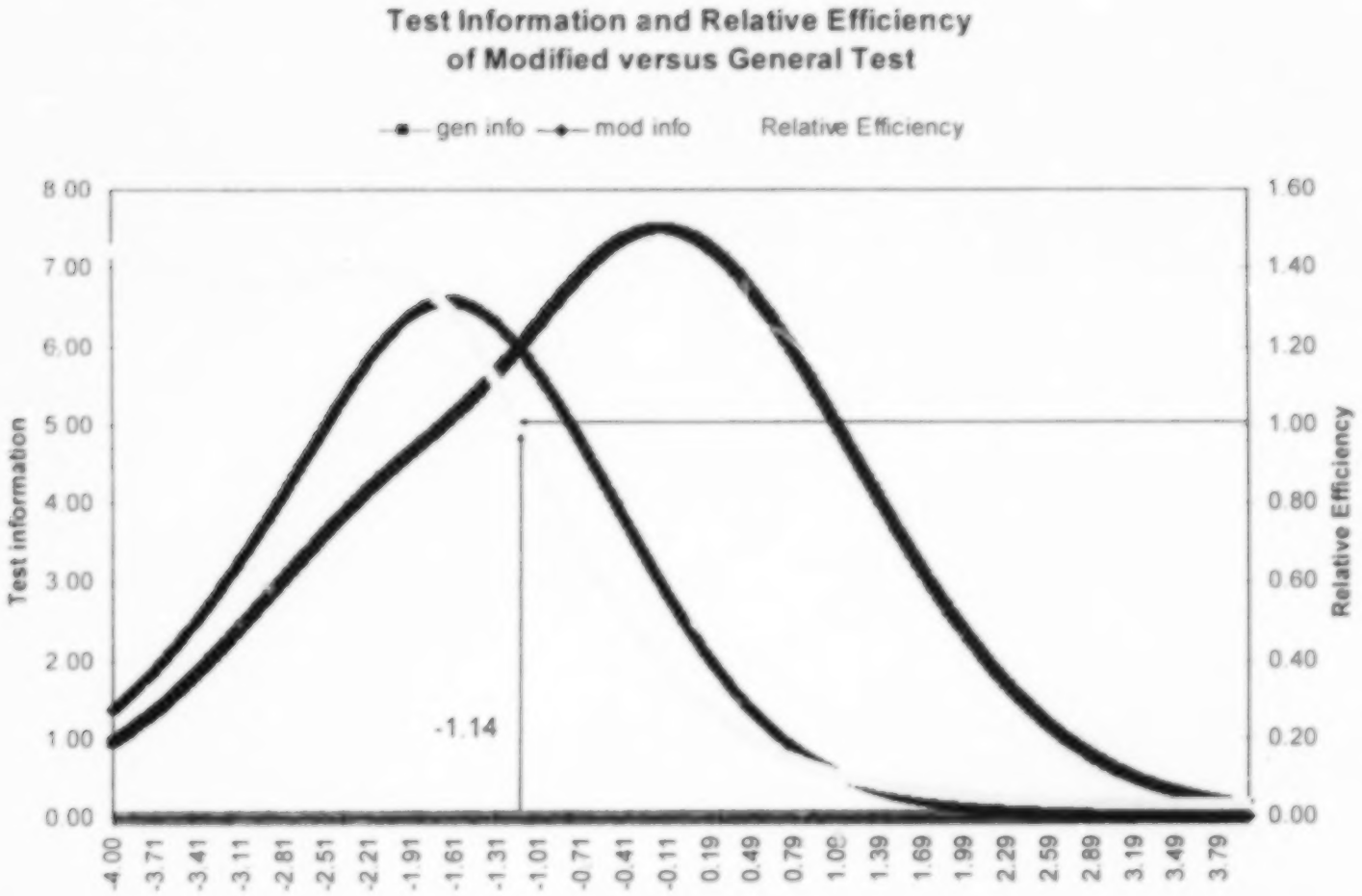

Figure 3. Test information by math ability level